Journal of Computational and Applied Mathematics 233 (2009) 36–45

Contents lists available at ScienceDirect

Journal of Computational and Applied Mathematics

journal homepage: www.elsevier.com/locate/cam

A fuzzy adaptive network approach to parameter estimation in cases where independent variables come from an exponential distribution

Turkan Erbay Dalkilic [a,*], Aysen Apaydin [b]

[a] *Karadeniz Technical University, Faculty of Arts and Sciences, Department of Statistics and Computer Sciences, 61080, Trabzon,Turkey*

[b] *Ankara University, Faculty of Sciences, Department of Statistics, Ankara, Turkey*

ARTICLE INFO

Article history:
Received 17 November 2007
Received in revised form 25 July 2008

Keywords:
Fuzzy adaptive network
Switching regression
Exponential distribution
Membership function
Validity criterion

ABSTRACT

In a regression analysis, it is assumed that the observations come from a single class in a data cluster and the simple functional relationship between the dependent and independent variables can be expressed using the general model; $Y = f(X)+\varepsilon$. However; a data cluster may consist of a combination of observations that have different distributions that are derived from different clusters. When faced with issues of estimating a regression model for fuzzy inputs that have been derived from different distributions, this regression model has been termed the 'switching regression model' and it is expressed with $Y^L = f^L(X) + \varepsilon^L$ $(L = \prod_{i=1}^{p} l_i)$. Here l_i indicates the class number of each independent variable and p is indicative of the number of independent variables [J.R. Jang, ANFIS: Adaptive-network-based fuzzy inference system, IEEE Transaction on Systems, Man and Cybernetics 23 (3) (1993) 665–685; M. Michel, Fuzzy clustering and switching regression models using ambiguity and distance rejects, Fuzzy Sets and Systems 122 (2001) 363–399; E.Q. Richard, A new approach to estimating switching regressions, Journal of the American Statistical Association 67 (338) (1972) 306–310].

In this study, adaptive networks have been used to construct a model that has been formed by gathering obtained models. There are methods that suggest the class numbers of independent variables heuristically. Alternatively, in defining the optimal class number of independent variables, the use of suggested validity criterion for fuzzy clustering has been aimed. In the case that independent variables have an exponential distribution, an algorithm has been suggested for defining the unknown parameter of the switching regression model and for obtaining the estimated values after obtaining an optimal membership function, which is suitable for exponential distribution.

1. Introduction

In many applied studies, problems arise from independent variables in a data cluster that are not from the same distribution, or even if they come from the same distribution, they do not share similar distribution parameters. Under this condition, when one single regression model belongs to the set of data, the predictions, which will be concluded from this model, will likely have high errors. When the data cluster has a structure similar to that mentioned above, it forms a regression model for each different class and can reach the results prediction with the combination of the predictions that are derived from these models. An appropriate model for this method is called the 'switching regression model' [16,17]. The process of setting up the switching regression model includes data clustering procedures, setting up more than one

* Corresponding author. +90 462 377 35 61; fax: +90 462 325 31 95.
E-mail addresses: tedalkilic@gmail.com (T.E. Dalkilic), aapaydin@ankara.edu.tr (A. Apaydin).

0377-0427/$ – see front matter © 2008 Elsevier B.V. All rights reserved.
doi:10.1016/j.cam.2008.07.057

model and combining the results for the purpose of retrieving data. In early studies using the switching regression model, two appropriate sub-models were formed with the idea that the described data were derived from the two different classes rather than one single class. The studies that followed aimed at making, the switching regression model generalizable, with the notion that the data could be formed from the combination of observations from more than two classes. Hathaway and Bezdek [9], emphasized the fuzzy clustering and switching regression model.

In the case that, the class numbers of the data and the number of the independent variables are more than two, simultaneously the number of sub-models are increased. At this stage, the method attempts to utilize the neural networks, which are intended to solve complex problems and systems. When faced with issues in which the data belong to an indefinite or fuzzy class, the neural network, termed the adaptive network, is used for establishing the switching regression model.

Different studies examining fuzzy clustering and validity criteria exist in the literature. In a study by Xie, X. L. and Beni, G.; a validity criterion is suggested for fuzzy clustering [19]. In a study by Chen, M. S., and Wang, S. W., an analysis of fuzzy clustering was completed to determine fuzzy memberships, and in that study, a method was suggested for indicating the optimal class numbers that belong to the variables [5]. Bezdek, J. C., also presented an important study on fuzzy clustering [1].

There are many studies on the use of the adaptive network for parameter prediction. In a study by Chi-Bin, C. and Lee, E. S. a fuzzy adaptive network approach was established for fuzzy regression analysis [4] and it was studied on both fuzzy adaptive networks and the switching regression model [3]. Jang, J. R. studied the adaptive networks based on a fuzzy inference system [15]. In a study of Takagi, T. and Sugeno, M., the method for identifying a system using it's input-output data was presented [18]. James, P. D. and Donalt, W., were studied fuzzy regression using neural networks citejam. In a study by Cichocki, A. and Unbehauen, R., the different neural networks for optimization were explained [6].

One of the important problems of the practical application of fuzzy cluster theory is the indication of the membership function. Civanlar, M. R. and Trussell, H. J. [7] advanced a method for indicating the membership function relative to fuzzy clusters with characteristic qualities with a probability density function, of which the elements are known. In a study in [8] on membership functions, the different membership functions were described and some information was provided about necessary properties for establishing membership functions and mathematical forms of membership functions.

In this work, in defining the optimal class number of independent variables, the use of suggested validity criteria for fuzzy clustering has been applied. In the case of independent variables with exponential distributions, another aim of this work was to utilize an algorithm that has been suggested to define the unknown parameter of the switching regression model and to obtain the estimated values after obtaining an optimal membership function, which is suitable for exponential distribution.

The remainder of the paper is organized as follows. Section 2 explores the formation of membership functions appropriate for a distribution of independent variables that are exponentially distributed within the parameter prediction. In Section 3, the fuzzy if-then rules and the use of these rules will be introduced using adaptive networks for analysis. The determination of the optimal cluster numbers for the independent variables, using validity criteria will be advanced in the same section. Additionally, an algorithm is suggested for the parameter prediction of the switching regression model on the condition that the independent variables are exponentially distributed. A numerical application examining the work and validity of the suggested algorithm as well as a comparison of the algorithm with classical methods is provided in Section 4 where data are assumed to be claim sizes for an insurance company. In Section 5, a discussion and conclusion are provided.

Regarding the numerical application of claim sizes in Section 4, it is useful to note that the switching regression models can be used for actuarial risk management purposes about identifying risky portfolios and credibility regression applications [2].

2. Forming the membership function for an exponential distribution

If the problem in question is going to be solved with fuzzy clustering, the most important step is determining the membership function appropriate for the data cluster or clusters within the problem. To determine the membership function, there are membership functions based on intuitional descriptions: for special problems, there are membership functions that are based on reliability and there are theoretically based membership functions. In this section, the method suggested in [7] which can satisfy the theoretical need, based on the probability density function, will be used to form the membership function appropriate for the data cluster, which is derived from the exponential family. A membership function should provide the given conditions below to be an optimal membership function:

1. $E\{\mu(x)\,|x$ is distributed according to the underlying probability density function$\} \geqslant c$
2. $0 \leqslant \mu(x) \leqslant 1$
3. $\int \mu^2(x)\,\mathrm{d}(x)$ should be minimized. This condition is required to obtain a selective membership function.

Here, the confidence level c should be close to unity and E is the average membership value. Under these conditions the optimal membership function is given in the shape of

$$\mu(x) = \begin{cases} \lambda p(x) & \text{if } \lambda p(x) < 1 \\ 1 & \text{if } \lambda p(x) \geqslant 1. \end{cases} \tag{2.1}$$

Here, $p(x)$: probability density function
λ : constant [7].

In the given membership function, the form, $p(x)$, is determined as the probability density function related to the interested distribution. However, the fixed element λ, can be obtained by solving the problem, which is formed using conditions described for optimal membership function and given by Eq. (2.2),

$$\begin{aligned} &P : \mathrm{Min}_{\mu} f(\mu) = \frac{1}{2}\int_{-\infty}^{+\infty} \mu^2(x)\mathrm{d}(x) \\ &G(\mu) = c - E\{\mu\} = c - \int_{-\infty}^{+\infty} \mu(x)p(x)\mathrm{d}(x) \leqslant 0 \\ &\mu \in \Omega = \{\mu(x)\,|0 \leqslant \mu(x) \leqslant 1\}. \end{aligned} \tag{2.2}$$

The problem given with P can be solved with the Lagrange Multipliers Method to obtain the fixed element, λ. For this, the Lagrange Function is written,

$$L(\mu, \lambda) = \frac{1}{2}\int_{-\infty}^{+\infty} \mu^2(x)\mathrm{d}(x) + \lambda \left\{c - \int_{-\infty}^{+\infty} \mu(x)p(x)\mathrm{d}(x)\right\}. \tag{2.3}$$

Here, $\lambda \geqslant 0$ and Lagrange multiplier and c is a constant and $c < 1$.

When the membership function values provided in Eq. (2.1) are integrated into Eq. (2.3), the following form of the Lagrangian is obtained

$$L(\mu^*, \lambda) = \frac{1}{2}\int_{-\infty}^{+\infty} \{I\,(\lambda p(x))\,(\lambda p(x) - 1)^2 - \lambda^2 p^2(x)\}\mathrm{d}(x) + \lambda c \tag{2.4}$$

where

$$I(x) = \begin{Bmatrix} 0 & \text{if } x \leqslant 1 \\ 1 & \text{if otherwise} \end{Bmatrix}.$$

By using the obtained lagrange function, and by incorporating the values $I\,(\lambda p(x))$, the function with the integrated constant λ can be obtained in the shape of

$$L = -\frac{1}{2}\int_{-\infty}^{+\infty} \lambda^2 p^2(x)\mathrm{d}(x) + \lambda c.$$

When this function's derivative is taken with respect to the constant element λ and is equalized to zero, the equation

$$\frac{\partial L}{\partial \lambda} = -\lambda \int_{-\infty}^{+\infty} p^2(x)\mathrm{d}(x) + c = 0 \tag{2.5}$$

is obtained. With the solution of the obtained Eq. (2.5), the

$$\lambda = \frac{c}{\int_{-\infty}^{+\infty} p^2(x)\mathrm{d}(x)} \tag{2.6}$$

equation is gained and this constant element for the exponential distribution when the probability density function;

$$p(x) = \frac{1}{\nu}\mathrm{e}^{-\frac{x}{\nu}} \quad x > 0\ \nu > 0$$

related to exponential distribution is placed in Eq. (2.6), it is obtained in the shape of

$$\lambda = 2\nu c. \tag{2.7}$$

From Eq. (2.7) the general membership function provided by Eq. (2.1) is obtained, as follows, for exponential distribution

$$\mu(x) = 2c\mathrm{e}^{-\frac{x}{\nu}}. \tag{2.8}$$

Here c is a constant element and ν is a distribution parameter, which is considered a priori parameter.

In the data set derived from the exponential distribution, the limit of the data belonging to the cluster with one membership degree is dependent on the fixed element c and the parameter ν, which indicates the distribution. This limit, given with $a(c)$, is described by,

$$a(c) = \max\{0, -\nu \ln(2(1-c))\}. \tag{2.9}$$

The membership degrees of the observations with values greater than the $a(c)$ limit, will be calculated using the membership function. For the different values of c, the limit given with $a(c)$ should be indicated. By providing different values between 0 and 1 to c, the values that $a(c)$ can integrate are calculated from Eq. (2.9) in the form of

$$\begin{aligned} &-\nu \ln(2(1-c)) < 0 \quad \text{for } 0 < c < 0.5 \Rightarrow a(c) = 0 \\ &-\nu \ln(2(1-c)) = 0 \quad \text{for } c = 0.5 \Rightarrow a(c) = 0 \\ &-\nu \ln(2(1-c)) > 0 \quad \text{for } 0.5 < c < 1 \Rightarrow a(c) = -\nu \ln(2(1-c)). \end{aligned}$$

For the values of c less than 0.5, $a(c)$ is calculated as zero. Since the fixed element c is less than 1, for the exponential distribution, it is concluded that c can take a value in the interval of $0.5 \leqslant c < 1$. For the exponential distribution with ν parameters, it is indicated that the observations will integrate a membership degree of 1 into the interval $[0, a(c)]$. As a result, the optimal membership function for the exponential distribution function is obtained in the shape of

$$\mu(x) = \begin{cases} 2\,c\,\mathrm{e}^{-\frac{x}{\nu}} & \text{if } x > a(c) \\ 1 & \text{if } x \leqslant a(c). \end{cases} \tag{2.10}$$

To reach the most appropriate c value in order to form the membership function given by Eq. (2.10), the program formed in the MATLAB. This function is related to the data sets that are generated from the exponential distribution and are indicated by the membership functions related to the observations obtained with this membership function. When the different c values are used for the variables, of which the membership degree will be one, the $a(c)$ limit will also be different. With the generated program, the membership functions which equal the values for many different data sets that c can take are obtained and the most appropriate membership function is reached on the condition that $c = 0.6$.

3. Parameter estimation by adaptive network in case where fuzzy independent variables come from an exponential distribution

In forming a switching regression model, one of the most important points is that it is necessary to determine how many clusters, and therefore, the number of fuzzy sets that the data set should be divided into related independent variables. At this stage, there is problem of determining optimal numbers of clusters. In parameter prediction studies conducted via adaptive networks, the class numbers of data sets related to independent variables are determined intuitively initially. In this study, we aimed to use validity criterion to determination the optimal class number. Validity criterion has been designed to identify the optimal numbers of clusters and determine compacted and separated clusters. There are a number of significant validity criteria for clusters in literature. In this study, the S function also called the Xie-Beni (1991) index will be used [19].

The adaptive network used to predict the unknown parameters of regression model is based on fuzzy if-then rules and fuzzy inference system. When issues of estimating a regression model to fuzzy inputs from different distributions arose, the Sugeno Fuzzy Inference System is appropriate and the proposed fuzzy rule in this case is indicated as

$$R^L = \text{ If; } (x_1 = F_1^L \text{ and } x_2 = F_2^L \text{ and } \ldots x_p = F_p^L).$$

Then; $Y = Y^L = c_0^L + c_1^L x_1 + \cdots + c_p^L x_p$.

Here F_i^L stands for fuzzy cluster and Y^L stands for system output according to R^L rule [15,18].

The weighted mean of the models obtained according to fuzzy rules is the output of Sugeno Fuzzy Inference System and a common regression model for data from different classes is indicated with this weighted mean. Neural networks that enable the use of fuzzy inference systems for fuzzy regression analysis is known as an adaptive network. Neural networks are used to obtain a good approach to regression functions and were formed via neural and adaptive network connections consisting of five layers [10–12,14].

Neurons, which form the network are characterized with parameter functions. In the process of adaptive networks, that consist of five layers, functional relationships between dependent and independent variables are modelled and predictions are obtained from these models [3,4].

Fuzzy rule numbers of the system depend on the numbers of independent variables and class or fuzzy sets number forming independent variables. When independent variable numbers are indicated with p and if the fuzzy class number associated with each variable is indicated by l_i $(i = 1, \ldots, p)$, the fuzzy rule number is indicated by

$$L = \prod_{i=1}^{p} l_i. \tag{3.1}$$

The aim of the fuzzy adaptive network is to obtain the model of the relationship between the input-output data couples. The difference between the output (prediction) obtained from the network related to the model and the output, which is targeted, is described as an error measure. The network that will be used, should be disciplined to perform this error minimally. When this error measure is smaller than the previously determined acceptable small error, the discipline of the network will terminate.

Parameter prediction with adaptive networks is based on the principle of the minimum error criterion. Different kinds of algorithms are proposed in the literature for forming regression models associated with data derived from different classes, as well as the process of gaining collective prediction sets based on this regression model. There are two significant steps in the process of prediction. The first step includes determining an a priori parameter set characterizing class from which the data are derived, and updating these parameters within the process. The other is to determine a posteriori parameters associated with regression models. In the proposed algorithm, a posteriori parameter sets $c_i^L = (a_i^L, b_i^L)$ are obtained by solving the linear programming problem, which is suggested in [12,13]. The problem is indicated as

$$\begin{aligned}
&\min \sum_{k=1}^{N} \sum_{L=1}^{m} \sum_{i=0}^{p} \bar{w}^L b_i^L x_{ik} \quad \left(= \min \sum_{k=1}^{N} \hat{e}_k \right) \\
&b_i^L \geqslant 0 \quad i = 1, \ldots, p\; L = 1, \ldots, m \\
&\sum_{L=1}^{m} \sum_{i=0}^{p} \bar{w}^L a_i^L x_{ik} + (1-\alpha) \sum_{L=1}^{m} \sum_{i=0}^{p} \bar{w}^L b_i^L x_{ik} \geqslant y_k + (1-\alpha) e_k \\
&-\sum_{L=1}^{m} \sum_{i=0}^{p} \bar{w}^L a_i^L x_{ik} + (1-\alpha) \sum_{L=1}^{m} \sum_{i=0}^{p} \bar{w}^L b_i^L x_{ik} \geqslant -y_k + (1-\alpha) e_k.
\end{aligned} \tag{3.2}$$

Updating a priori parameters initially determined intuitively within the process is based on back propagation errors.

The process of determining parameters for the switching regression model begins with determining class numbers of independent variables and a priori parameters. In this study, it is aimed to use validity criterion based on fuzzy clustering as an alternative to intuitive methods in determining class numbers. Moreover, a structure is formed that includes determining a priori parameters as dependents to alter the data set intervals. When updating a priori parameters, instead of the method in which spread back error bringing a series of transaction, a process has been formed to spread the back error bringing a series of transactions, a process has been formed which enables to review all the values that parameter might have the lowest error. The algorithm associated with the proposed method of determining switching regression models in the case of independent variables derived from an exponential distribution, is proposed as follows.

3.1. Proposed Algorithm

Step 0: Optimal class numbers related to the data set associated with the independent variables are determined. Optimal value of class number l_i, ($l_i = 2, l_i = 3 \ldots l_i = \max$) can be obtained by minimizing the fuzzy clustering validity function S_i. This function is expressed by

$$S_i = \frac{\frac{1}{n} \sum_{i=1}^{l_i} \sum_{j=1}^{n} \mu_{ij}^m \left\| \nu_i - x_j \right\|^2}{\min_{i \neq j} \left\| \nu_i - \nu_j \right\|^2} = \frac{comp}{sep}. \tag{3.3}$$

As it can be seen in this statement, cluster centers, which are well-separated produce a high value of separation such that a smaller S_i value is obtained. When the lowest S_i value is observed, class number (l_i) with the lowest S_i value is defined as an optimal class number.

Step 1: An a priori parameter set is determined. The a priori parameters that display the cluster centers, which belong to the independent variables depend on the span and fuzzy class numbers of the independent variables. This is indicated by

$$\nu_i = \min(X_i) + \frac{\max(X_i) - \min(X_i)}{(l_i - 1)} (i - 1) \quad i = 1, \ldots, p. \tag{3.4}$$

Step 2: $\bar{w}^L$ weights are counted, which are used to form matrix B to be used in counting posteriori parameter set. When the exponential distribution function, which has the parameter set of $\{\nu_i\}$, and the membership function, which will be used in the calculation of these sets are regarded, membership functions are described in the shape of

$$\mu_{F_h}(x_{ij}) = \begin{cases} 2c\mathrm{e}^{-\frac{x_{ij}}{\nu_i}} & \text{if } x_{ij} > a(c)_i \\ 1 & \text{if } x_{ij} \leqslant a(c)_i. \end{cases} \tag{3.5}$$

Here, the set of $\{\nu_i\}$ shows the a priori parameters. The $\bar{w}^L$sets are the normalization of the rule's weights, which is indicated with w^L and this is calculated using

$$\bar{w}^L = \frac{w^L}{\sum_{L=1}^{m} w^L}. \tag{3.6}$$

Step 3: On the condition that the independent variables are fuzzy and the dependent variables are crisp, a posteriori parameter set $c_i^L = \left(a_i^L, b_i^L\right)$ is obtained as crisp numbers in the shape of, $c_i^L = a_i^L$. In that condition,

$$Z = (B^T B)^{-1} B^T Y$$

equation is used to determine the a posteriori parameter set. Here B, Y and Z defined as

$$B = \begin{bmatrix} \bar{w}_1^1, & \dots, & \bar{w}_1^L, & \bar{w}_1^1 x_{11}, & \dots, & \bar{w}_1^L x_{11}, & \dots, & \bar{w}_1^1 x_{p1}, & \dots, & \bar{w}_1^L x_{p1} \\ \vdots & & & & & \bar{w}_k^l x_{jk} & & & & \vdots \\ \bar{w}_n^1 & \dots, & \bar{w}_n^L, & \bar{w}_n^1 x_{1n}, & \dots, & \bar{w}_n^L x_{1n}, & \dots, & \bar{w}_n^1 x_{pn}, & \dots, & \bar{w}_n^L x_{pn} \end{bmatrix},$$

$Y = [y_1, y_2, \dots, y_n]^T$ and $Z = \left[a_0^1, \dots, a_0^L, a_1^1, \dots, a_1^L, a_p^1, \dots, a_p^L\right]^T$.

Step 4: By using a posteriori parameter set $c_i^L = \left(a_i^L\right)$ obtained in Step 3, the switching regression model indicated by

$$Y^L = c_0^L + c_1^L x_1 + c_2^L x_2 + \dots + c_p^L x_p$$

are constituted. Setting out from the models and weights specified in step 2, the prediction values are obtained using

$$\hat{Y} = \sum_{L=1}^{m} \bar{w}^L Y^L. \tag{3.7}$$

Step 5: Error related to the model is counted as

$$\varepsilon = \frac{1}{N} \sum_{k=1}^{N} \varepsilon_k^2 = \frac{1}{N} \sum_{k=1}^{N} \left(y_k - \hat{y}_k\right)^2. \tag{3.8}$$

If $\varepsilon < \phi$, then a posteriori parameter has been obtained as parameters of regression models to be formed, the process is determined.

If $\varepsilon \geqslant \phi$, then, step 6 begins,

Here ϕ, is a law stable value determined by decision maker,

Step 6: Central priori parameters specified in Step l, are updated with

$$v_i' = v_i \pm t \tag{3.9}$$

in a manner that it increases from the lowest value to the highest and decreases from the highest value to the lowest. Here, t is size of step;

$$t = \frac{\max(x_{ji}) - \min(x_{ji})}{a} \quad j = 1, \dots, n \; i = 1, \dots, p$$

and a is stable value, which is a determinant of step size and therefore iteration number.

Step 7: Predictions for each a priori parameter obtained by change and error criteria related to these predictions are counted with

$$\varepsilon_k = Y_k \{-\} \hat{Y}_k. \tag{3.10}$$

Here;

Y_k : k. predicted outcome

$\hat{Y}_k$: k. network output of input vector

$\{-\}$: Difference operator.

The lowest error criterion is defined. A priori parameters with the lowest error specified, and predictions obtained via models related to these parameters is considered to be the output. This method can also be used when the dependent variable is fuzzy.

In the proposed algorithm, determining the coefficients of the fuzzy rules, the equation $Z = (B^T B)^{-1} B^T Y$ is used which is located in Step 3. Also, the optimization problem which is located in Eq. (3.2) can be used. In the proposed method, linear models with real number coefficients are obtained. The fuzzy inference system, which is used in the proposed algorithm determines a real-valued $F^L(X)$ of p input variables by the following process,

$$X \longrightarrow \textit{fuzzifier} \longrightarrow \textit{fuzzyinference} \longrightarrow \textit{defuzzifier} \longrightarrow \hat{Y} \quad .$$

Where $X = [x_1, x_2, \dots, x_p]$ is a crisp multivariate input vector and $\hat{Y}$ is a crisp univariate output. The degree to which the input variables belong to fuzzy rules is determined by fuzzy clustering. A given input x_i is applied to all L rules simultaneously. An input can activate two or more rules depending on how well it matches the input fuzzy sets of each rule. If more than two rules are activated, their outputs are combined to produce the final output. The final crisp output $\hat{Y}$ represents combined evidence provided by each rule. The fuzzy process which has a tree step, is placed in the proposed algorithm in this manner. The first step, called fuzzification, is located in Step 2, where for a given input X, calculate rule-l weight as a fuzzy membership function, $\mu_{F_h}(x_{ij})$ via Eq. (3.5). The second step, called rule output evaluation, is performed in the same step by calculating the w^L which is obtained by multiplying the rules weight. Lastly, in the Step 4, the defuzzification step, according to Eq. (3.7), the individual rule outputs are additively combined.

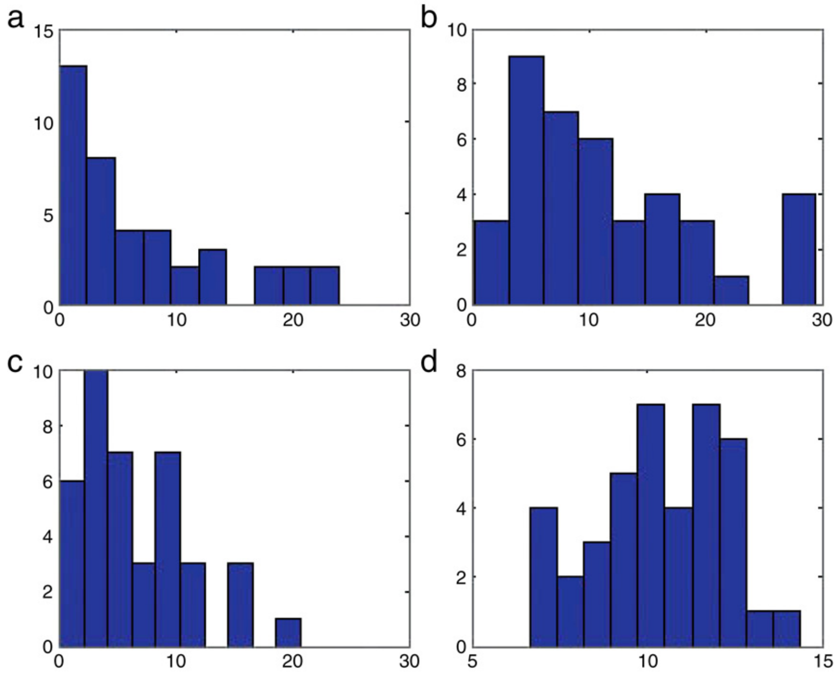

Fig. 4.1. Histograms of dependent and independent variables.

In the proposed algorithm, the Sugeno fuzzy inference system is used to form the fuzzy rule. In the studies of Sugeno, triangular, trapezoidal and piecewise-polynomial membership functions are used to form of the fuzzy rule generally. In this work, it is considered that the independent variables have exponential distribution and the membership function, which is appropriate for this distribution, is obtained. The use of this membership function in the forming of the fuzzy rule is provided to minimizing errors associated with the estimates. Using the program, which was formed in the MATLAB for proposed algorithm, the value of the prime parameters can change minimally, and the prediction errors can be calculated with the use of these values. As a result, the optimal value of prime parameters can selected by this program.

4. Numerical examples

The values related to the data set with three independent variables and one dependent variable are displayed in Table 4.1. The algorithm, proposed in the Section 3, with the switching regression model for this data set derived using simulation and the predictions associated with this model are obtained. The prediction values derived from the adaptive network, which is related to the data set and the errors related to these predictions, are displayed in Table 4.1. In addition, to show the performance evaluation of the proposed method, the prediction value from this method and the errors related to these predictions are compared with other predictions and the errors, which are obtained from the Least Square Method (LSM) and multi-layer feed forward neural network. This neural network is based on back-propagation algorithm. For the obtain prediction value, the neural networks toolbox of the MATLAB is used.

In the prediction process by multi-layer feed forward neural network, there are determinations of the two important parameters. One of them is learning rate the other one is number of epochs. The training stops if the number of iterations exceeds epochs and the learning rate is multiplied times the negative of the gradient to determine the changes to the weights and biases. The larger the learning rate, the bigger the step. If the learning rate is made too large, the algorithm becomes unstable. If the learning rate is set too small, the algorithm takes a long time to converge [6,20].

In the prediction process by neural network with the back propagation algorithm (BPNN) a program is generated to determination of the optimal value of learning rate and epoch. And most appropriate prediction is reached on the condition that *learning rate* $= 0.05$ and *Epoch* $= 500$. The predictions that are obtained with these two methods (LSM and BPNN) and the errors related with these predictions are also displayed in Table 4.1. The errors belonging to each result which are calculated by Eq. (3.8) are given in the Table 4.1. Here, X_1, X_2, X_3 can be considered as total claim amounts in three different portfolios of policies and Y can be considered the total claim amount in an insurance set up.

The histograms of the independent and dependent variables are given in Fig. 4.1. In Fig. 4.1(a), (b) and (c), are histograms of independent variables X_1, X_2, X_3 and (d) is histogram of the dependent variable Y.

For every three variables, fuzzy class numbers are determined as $l = 2$. For the predictions that are obtained from the fuzzy inference rule that is formed according to this indicated fuzzy class numbers, the error value is calculated as $\varepsilon_{Network} = 0.5343$ by making use of Eq. (3.8). The error value related to the predictions, which are obtained from the LSM, is

Table 4.1
Predictions and Error Values for data set related to tree independent variables, which are derived from exponential distributions

X_1	X_2	X_3	Y	$\hat{y}_{LSM}$	$e_{(LSM)i}$	$\hat{y}_{Network}$	$e_{(Network)i}$	$\hat{y}_{BPNN}$	$e_{(BPNN)i}$
23.9866	27.3134	9.1045	9.1349	10.2232	−1.0884	9.3061	0.1713	−1.4395	10.5744
2.1028	0.2975	0.0992	6.6688	10.3982	−3.7294	6.6707	0.0018	−0.0775	6.7463
7.4684	17.1515	5.7172	10.2507	10.2075	0.0431	11.6100	1.3593	−0.3237	10.5744
5.3240	7.1580	2.3860	10.5754	10.3308	0.2445	10.5095	−0.0659	0.0020	10.5734
17.7532	26.7387	8.9129	7.7071	10.1702	−2.4632	9.2756	1.5686	−2.8673	10.5744
11.4871	10.4818	3.4939	12.3818	10.3434	2.0384	13.7676	1.3858	1.8075	10.5743
4.8773	18.5408	6.1803	12.3783	10.1620	2.2164	10.4505	−1.9279	1.8039	10.5744
0.1494	8.4027	2.8009	9.9247	10.2619	−0.3372	9.9262	0.0015	−0.6496	10.5743
13.7812	5.4494	1.8165	10.6546	10.4387	0.2159	10.6810	0.0265	0.3564	10.2982
4.7060	3.1544	1.0515	10.3493	10.3826	−0.0333	10.4433	0.0940	0.6267	9.7226
7.6451	3.2245	1.0748	9.6266	10.4105	−0.7839	9.5424	−0.0842	0.5519	9.0747
12.5593	17.1961	5.7320	11.4516	10.2570	1.1946	10.6929	−0.7587	0.8772	10.5744
20.3892	5.4095	1.8032	8.8234	10.5043	−1.6809	8.8770	0.0537	−0.1988	9.0222
10.7216	11.7026	3.9009	14.3664	10.3183	4.0481	12.6556	−1.7108	3.7920	10.5744
1.5513	2.4532	0.8177	9.7272	10.3617	−0.6345	9.7142	−0.0130	0.2039	9.5233
4.1631	17.9549	5.9850	10.2279	10.1634	0.0645	11.2398	1.0120	−0.3465	10.5744
21.9250	7.1312	2.3771	12.1335	10.4945	1.6390	11.9361	−0.1975	1.7347	10.3988
19.9021	29.4929	9.8310	10.1186	10.1516	−0.0330	8.7645	−1.3541	−0.4558	10.5744
4.2247	28.8268	9.6089	9.8087	10.0069	−0.1982	9.8940	0.0853	−0.7657	10.5744
17.9281	13.5049	4.5016	8.3353	10.3631	−2.0278	8.7128	0.3775	−2.2391	10.5744
0.2982	6.8628	6.1765	10.5888	10.3179	0.2709	10.3902	−0.1986	0.0187	10.5701
2.1760	23.0028	20.7025	7.3276	10.1791	−2.8514	7.8328	0.5052	−3.2468	10.5744
8.3877	17.2389	15.5150	11.4286	10.2964	1.1323	11.5850	0.1563	0.8542	10.5744
0.0496	10.3539	9.3185	13.2471	10.2814	2.9657	12.5959	−0.6512	2.6728	10.5743
0.7477	17.0361	15.3325	8.6164	10.2232	−1.6067	8.0483	−0.5682	−1.9580	10.5744
1.1330	10.7948	9.7153	11.7160	10.2878	1.4282	11.4093	−0.3067	1.1417	10.5743
1.1077	4.1851	3.7666	12.5080	10.3520	2.1560	12.2557	−0.2523	2.1735	10.3345
4.6291	3.4210	3.0789	6.8125	10.3941	−3.5815	6.4809	−0.3316	−2.2982	9.1107
1.5886	4.1733	3.7559	7.1181	10.3568	−3.2387	7.6227	0.5046	−3.1756	10.2937
1.1083	7.6374	6.8737	11.1423	10.3183	0.8240	11.6119	0.4696	0.5697	10.5726
0.0770	12.9870	11.6883	9.2002	10.2560	−1.0558	9.8125	0.6123	−1.3742	10.5744
6.8676	3.7004	3.3303	11.3800	10.4134	0.9666	11.7631	0.3831	2.5587	8.8213
2.9448	18.2323	16.4090	11.6312	10.2331	1.3981	10.9523	−0.6790	1.0568	10.5744
13.4276	8.3950	7.5555	11.4238	10.4321	0.9917	11.0266	−0.3972	0.8781	10.5457
3.1367	4.6269	4.1642	12.5805	10.3676	2.2129	12.2695	−0.3110	2.2466	10.3339
2.7120	12.1315	10.9183	11.3372	10.2903	1.0469	10.7756	−0.5617	0.7628	10.5744
9.3612	7.9092	7.1183	12.3817	10.3969	1.9848	12.3285	−0.0531	1.8435	10.5582
3.7238	5.8857	5.2971	7.5951	10.3611	−2.7660	7.5659	−0.0292	−2.9294	10.5245
1.1323	11.8603	10.6742	9.9604	10.2774	−0.3170	10.9473	0.9869	−0.6140	10.5744
5.5758	9.7104	8.7393	9.6866	10.3421	−0.6555	10.3828	0.6962	−0.8874	10.5740
ERROR				$\varepsilon_{LSM} = 3.3399$		$\varepsilon_{Network} = 0.5343$		$\varepsilon_{BPNN} = 2.8832$	

calculated as $\varepsilon_{LMS} = 3.3399$ and the error value obtained from the BPNN is calculated as $\varepsilon_{BPNN} = 2.8832$ by making use of the same equation. From the beginning step of the proposed algorithm, for each variable the class numbers are determined as two. The number of the fuzzy inference rules formed according to this indicated fuzzy class numbers are defined as

$$L = \prod_{i=1}^{p=3} l_i = l_1 \times l_2 \times l_3 = 2 \times 2 \times 2 = 8.$$

The models that are obtained from the eight fuzzy inference rules are,

$$\hat{y}_1 = 4553 + 3419x_1 + 39\,964x_2 - 44\,787x_3$$
$$\hat{y}_2 = -41\,751 - 600x_1x - 56\,859x_2 + 90\,856x_3$$
$$\hat{y}_3 = 28\,630 - 00\,057x_1 + 10\,374x_2 - 36\,001x_3$$
$$\hat{y}_4 = 81 + 3x_1 + 5x_2 - 10x_3$$
$$\hat{y}_5 = -876 + 66x_1 - 1816x_2 + 900x_3$$
$$\hat{y}_6 = 2842 - 12x_1 + 2367x_2 - 5346x_3$$
$$\hat{y}_7 = -542 - 4x_1 - 1165x_2 + 3616x_3$$
$$\hat{y}_8 = x_3.$$

The figures of errors obtained from three methods are given in Fig. 4.2. Fig. 4.2(a) shows the errors related to the predictions that are obtained from the proposed algorithm, (b) shows the errors related to the predictions that are obtained from The Least Squares Method, (c) shows the errors related to the predictions that are obtained from the multi layer neural network based back propagation algorithm and (d) shows the comparison of the errors.

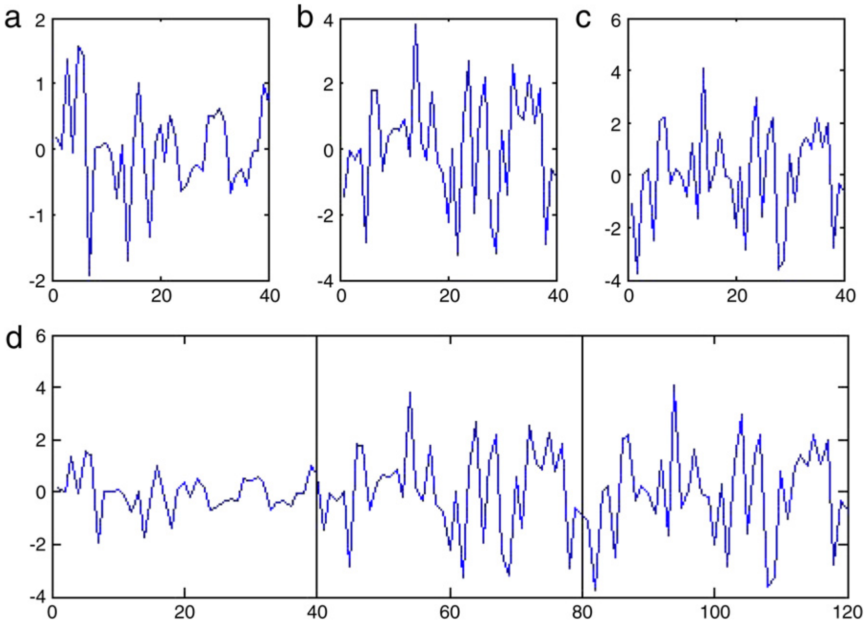

Fig. 4.2. Graphs for errors related to data set in Table 4.1.

Table 4.2
Beginning and result values related to the centers obtained from the network

	X_1		X_2		X_3	
	Class 1	Class 2	Class 1	Class 2	Class 1	Class 2
Initial v_i	0.0496	23.9866	0.2975	29.4929	0.0992	20.7025
Conclusion v_i	1.6689	22.3672	1.9169	27.8736	1.7186	19.0831

At the Table 4.2, from the proposed algorithm, the beginning center values that are indicated when models related to fuzzy rules and the center values where the best predictions are provided.

5. Conclusions

In recent years; many application areas have utilized the adaptive networks that occur under the heading of ‘neural networks’ and that give effective conclusions in the obtaining of the prediction values related to the data. In this regard; different methods have been introduced and several algorithms proposed according to these described methods. In the proposed algorithms, before starting the process of solving, the class number of the independent variables are determined heuristically and are bound to numbers, that are indicated at the beginning of the solution process. The fuzzy class numbers, which are determined intuitionally, are effective on a number of the models that are established in the solving process by being dependent to the variable numbers that occur in the model. The number of a priori and a posteriori parameters obtained in the solving process are dependent to the class number that is determined in the beginning.

In this study, the validity criterion based on the fuzzy clustering is utilized rather than the mere use of intuition for the determination of class numbers associated with independent variables. The criterion, which is used to determine the fuzzy class numbers, is more reliable than the intuitional determination, since it is based on calculations. With the method that is used to update a priori parameters, the procedure density of the algorithms are used before it is tried to be decreased.

The data sets handled in the studies are used to form the switching regression model. The independent variables are derived from the normal distribution, and regression models are formed using membership functions that are appropriate to the normal distribution. In this study, it is considered that the independent variables associated with problems in the application of data sets with exponential distribution and according to this, the membership function which is appropriate for the exponential distribution is obtained. After, the most appropriate parameters are determined for this membership function, which is obtained using simulation studies.

The network is formed that contains the determination of the a priori parameter sets and optimal class numbers related to data sets associated with independent variables are obtained. When obtaining membership functions appropriate for the distributions used in calculating the a posteriori parameter set, the formation of the switching regression models using a posteriori parameter set and obtaining prediction values from these formed models are performed. The difference between the obtained prediction values and the observed values, in other words, the network that decreases errors to a minimum level is formed based on the adaptive network architecture that includes fuzzy inference systems, which are based on fuzzy rules.

The prediction values obtained from networks that are appropriate to the algorithm proposed in the application stage and the prediction values obtained from the LSM and the prediction values obtained from the multi layer neural networks based back-propagation are compared. According to the indicated error criterion, the errors related to the predictions that are obtained from the network are less than errors obtained from the other methods.

The process followed in the proposed method can be accepted as ascendent from other methods since it doesn't let the intuitional predictions and it brings us to the smallest error. At the same time; this method has the robust feature since it doesn't affected by the contradictory observations that can occur at independent variables. In the subsequent studies, this method can be compared with other robust methods. This method's effectiveness can be examined when the contradictory observation is seen in the dependent variable.

Acknowledgments

The authors would like to thank the referees for very useful comments and suggestions.

References

[1] C.J. Bezdek, R. Ehrlich, W. Full, FCM: The fuzzy c_means clustering algorithm, Computer and Geoscience 10 (2-3) (1984) 191–203.
[2] H. Bühlmann, A. Gisler, A Course in Credibility Theory and its Applications, Springer, 2005.
[3] C. Chi-Bin, E.S. Lee, Switching regression analysis by fuzzy adaptive network, European Journal of Operational Research 128 (2001) 647–663.
[4] C. Chi-Bin, E.S. Lee, Applying fuzzy adaptive network to fuzzy regression analysis, An International Journal Computers & Mathematics With Applications 38 (1998) 123–140.
[5] M.S. Chen, S.W. Wang, Fuzzy clustering analysis for optimizing fuzzy membership functions, Fuzzy Sets and Systems 103 (1999) 239–254.
[6] A. Cichocki, R. Unbehauen, Neural Networks for Optimization and Signal processing, John Wiley &Sons, New York, 1993, p. 522.
[7] M.R. Civanlar, H.J. Trussell, Constructing membership functions using statistical data, Fuzzy Sets and Systems 18 (1986) 1–13.
[8] J. Dombi, Membership functions as an evaluation, Fuzzy Sets and Systems 35 (1990) 1–21.
[9] R.J. Hathaway, J.C. Bezdek, Switching regression models and fuzzy clustering, IEEE Transactions on Fuzzy Systems 1 (3) (1993) 195–204.
[10] S. Horikawa, T. Furuhashi, Y. Uchikawa, On fuzzy modeling using fuzzy neural networks with the back-propagation algorithm, IEEE Transections on Neural Networks 3 (5) (1992) 801–806.
[11] H. Ishibuchi, M. Nei, Fuzzy regression using asymmetric fuzzy coefficients and fuzzied neural networks, Fuzzy Sets and Systems 119 (2001) 273–290.
[12] H. Ishibuchi, H. Tanaka, Fuzzy regression analysis using neural networks, Fuzzy Sets and Systems 50 (1992) 257–265.
[13] H. Ishibuchi, H. Tanaka, H. Okada, An architecture of neural networks with interval weights and its application to fuzzy regression analysis, Fuzzy Sets and Systems 57 (1993) 27–39.
[14] P.D. James, W. Donalt, Fuzzy regression by fuzzy number neural networks, Fuzzy Sets and Systems 112 (2000) 371–380.
[15] J.R. Jang, ANFIS: Adaptive-network-based fuzzy inference system, IEEE Transactions on Systems, Man and Cybernetics 23 (3) (1993) 665–685.
[16] M. Michel, Fuzzy clustering and switching regression models using ambiguity and distance rejects, Fuzzy Sets and Systems 122 (2001) 363–399.
[17] E.Q. Richard, A new approach to estimating switching regressions, Journal of the American Statistical Association 67 (338) (1972) 306–310.
[18] T. Takagi, M. Sugeno, Fuzzy identification of systems and its applications to modeling and control, IEEE Transactions on Systems, Man and Cybernetics 15 (1) (1985) 116–132.
[19] X.L. Xie, G. Beni, A validity measure for fuzzy clustering, IEEE Transactions on Pattern Analysis and Machine Intelligence 13 (8) (1991) 841–847.
[20] J.M. Zaruda, Introduction to Artificial Neural Systems, West Publishing Company, St. Paul, 1992, p. 683.